

## Discrete Human Dihydrofolate Reductase Gene Transcripts Present in Polysomal RNA Map with Their 5' Ends Several Hundred Nucleotides Upstream of the Main mRNA Start Site

JEFFREY N. MASTERS<sup>†</sup> AND GIUSEPPE ATTARDI\*

*Division of Biology, California Institute of Technology, Pasadena, California 91125*

Received 24 September 1984/Accepted 13 December 1984

The 5' ends of dihydrofolate reductase (DHFR)-specific transcripts have been mapped in the 5'-flanking region of the amplified DHFR gene of the human methotrexate-resistant cell line 6A3 by primer extension and S1 protection experiments. The main 5' end, at position -71 relative to the first nucleotide of the DHFR reading frame, corresponds to the recently identified main transcription initiation site for the DHFR gene and pertains to transcripts representing approximately 99% of the DHFR-specific polysomal polyadenylic acid-containing RNA, and including the previously described DHFR mRNAs with sizes of 3.8, 1.0, and 0.8 kilobases. At least six other minor 5' ends have been mapped to nucleotide positions -449 to -480 upstream of the DHFR gene and pertain to approximately 1% of the DHFR-specific polysomal polyadenylic acid-containing RNA. These upstream initiating transcripts appear to include five major discrete species with sizes of 4.3, 3.8, 3.1, 2.1, and 1.0 kilobases and four minor ones with sizes of 7.3, 5.0, 1.4, and 0.8 kilobases. These species, with the exception of those of 3.1- and 2.1-kilobase sizes, also have been found in VA<sub>2</sub>-B cells, the parental line of 6A3, and in HeLa cells. The upstream initiating transcripts present in all three cell lines are increased in amount in 6A3 cells as compared with the other cell lines, in about the same proportion as the three identified DHFR mRNAs.

The analysis of the mode of expression of the dihydrofolate reductase (DHFR) gene in mammalian cells, which has been greatly facilitated by the availability of cell lines with amplified genes, has revealed an unsuspected complexity both in these and the parental cell lines, in the form of the existence of multiple species of DHFR mRNAs. Thus, four major polyadenylated [poly(A)<sup>+</sup>] species of DHFR mRNA in mouse cells (32), three species in Chinese hamster cells (22), and three main species in human cells (27) have been identified. These multiple mRNA species differ mainly in the length of the 3'-untranslated tail (10, 25, 32), which can be up to severalfold longer than the reading frame, as is the case for the major human DHFR mRNA (3.8-kilobase [kb] mRNA). Furthermore, there is good evidence indicating that the 3'-end tails of the multiple mRNAs are colinear (25, 32). These observations would be compatible with the idea that the same transcript could produce different forms of mRNA by processing and polyadenylation at alternative sites.

In the present work, an extensive mapping study of the 5' ends of DHFR-specific transcripts in human cell lines with an amplified or normal DHFR gene complement has unexpectedly revealed the presence, besides a main transcription initiation point at position -71 relative to the first nucleotide (nt) of the DHFR reading frame (11), of at least six additional minor sites several hundred nt's upstream of the DHFR gene, which correspond to 5' ends of DHFR-specific transcripts. These upstream initiating transcripts are found both in polysomal RNA, where they include several discrete species in the size range from 1.0 to 7.3 kb, and in nuclear RNA.

### MATERIALS AND METHODS

**Cell lines and methods of growth.** HeLa S3 cells were grown in suspension as described previously (3). VA<sub>2</sub>-B, an 8-azaguanine-resistant derivative (34) of the simian virus 40 transformed line WI-18-VA<sub>2</sub> (29), and its methotrexate-resistant variant 6A3 (24) were grown as described previously (27).

**RNA preparation and fractionation.** Polysomal RNA was prepared from HeLa, VA<sub>2</sub>-B, and 6A3 cells (27), and the poly(A)<sup>+</sup> fraction was separated from the non-poly(A)<sup>+</sup> fraction by oligodeoxythymidylic acid-cellulose chromatography (4). For the preparation of nuclear RNA, nuclei were isolated from 6A3 cells by a hypotonic shock procedure (20), and RNA was extracted by the urea-LiCl method (5).

**DNA preparation and labeling.** Plasmid DNAs were prepared by alkaline lysis (7), followed by CsCl-ethidium bromide centrifugation. For production of specific primers or fragments to be used in S1 protection experiments, 10 to 50 µg of plasmid DNA was first digested with a restriction enzyme producing the desired end, treated with calf intestinal alkaline phosphatase (Boeringer Mannheim Biochemicals), and labeled at the 5' ends with T4 polynucleotide kinase (Bethesda Research Laboratories) and [ $\gamma$ -<sup>32</sup>P]ATP (ICN Pharmaceuticals Inc.); the DNA was then spermine precipitated (19) to remove the bulk of unincorporated nt's and digested with a second restriction enzyme, and the fragments of interest, labeled at only one end, were isolated by elution from polyacrylamide gels (26). When the first restriction digest contained many fragments, the fragment of interest was isolated from the rest by elution from an agarose (36) or polyacrylamide (26) gel. DNA was sequenced by the method of Maxam and Gilbert (26).

**Primer extension and S1 protection experiments.** RNA-DNA hybridizations in both primer extension and S1 protection experiments were carried out in 10 µl of 0.1 M NaCl-0.02 M Tris (pH 8.0)-0.001 M EDTA at 60 to 75°C for

\* Corresponding author.

<sup>†</sup> Present address: Andrus Gerontology Center, University of Southern California, Los Angeles, CA 90007.

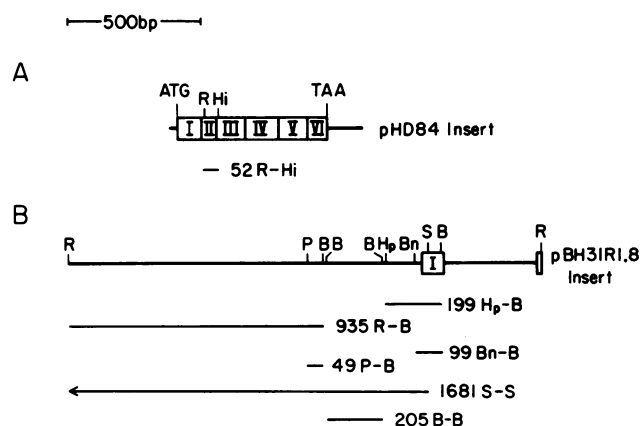


FIG. 1. Restriction fragments used in the primer extension, S1 protection, and RNA transfer hybridization experiments. (A) pH84 cDNA insert: the DHFR coding regions contained in the six exons are shown. The restriction sites utilized to isolate the 52-bp fragment (52R-Hi) used for primer extension analysis are indicated, and the fragment is aligned below. (B) The pBH31R1.8 insert is diagrammed to show the DHFR coding regions that it contains, the restriction sites used to generate the primers used in the extension experiments (99 Bn-B, 49 P-B), and the fragments used in the S1 protection experiments (199 Hp-B, 935 R-B); these fragments are aligned below. For the purpose of the above-mentioned experiments, each fragment was labeled with  $^{32}\text{P}$  at the 5' end of the non-sense strand. The fragments 205 B-B and 1681 S-S were  $^{32}\text{P}$ -labeled by nick translation for use in RNA transfer hybridization experiments. The arrow at the left end of 1681 S-S indicates that 328 nt's of pBR325 sequences are contained in this fragment, terminating at the nt at position 5112 of pBR325 (30). Restriction site designations: B, *Bst*NI; Bn, *Ban*II; Hi, *Hin*II; Hp, *Hpa*II; P, *Pvu*II; R, *Eco*RI; S, *Sfa*NI.

1 h, with the desired DNA fragment from 0.2  $\mu\text{g}$  of plasmid DNA and 1  $\mu\text{g}$  of polysomal poly(A) $^{+}$  RNA from 6A3 cells, and stopped by quick cooling in ice water. The primer extension reactions were performed in 50  $\mu\text{l}$  of 0.08 M Tris (pH 8.3)–0.05 M KCl–0.01 M MgCl $_2$ –0.01 M dithiothreitol–4 mM sodium pyrophosphate–20 U of RNasin (Promega Biotech), in the presence of 10 U of AMV reverse transcriptase (Life Science), at 42°C for 30 min, and were followed by phenol extraction, ethanol precipitation, and electrophoresis through 20-cm 7 M urea–5% polyacrylamide gels or 80-cm 8.3 M urea–5% polyacrylamide gels. The S1 nuclease digestions were carried out as specified in the legends of the figures.

**RNA transfer hybridization experiments.** Samples of polysomal poly(A) $^{+}$  RNA (1 or 2  $\mu\text{g}$ ) were electrophoresed through a 1% agarose slab gel containing 10 mM CH $_3$ HgOH in 1 $\times$  borate buffer (E buffer of Bailey and Davidson [6] containing 0.1 mM EDTA) and transferred to a Zeta Probe filter (Bio-Rad Laboratories) by electrophoresis in 1/3 $\times$  borate buffer as described previously (18). After baking the filter for 3 h at 80°C, the hybridization was carried out essentially as previously described (25).

## RESULTS

**Multiple 5' ends of human DHFR-specific RNAs.** Figure 1 shows the DHFR-specific cDNA and genomic restriction fragments used in the 5'-end mapping experiments, and Fig. 2 shows the experimental strategies used. The first primer extension reactions were carried out with 6A3 polysomal poly(A) $^{+}$  RNA as a template and the 52-nt *Eco*RI–*Hin*II fragment (Fig. 1A) derived from pH84 (25, 27) as a primer

(Fig. 2A). When the products of this reaction were electrophoresed on a denaturing polyacrylamide gel, a major band corresponding to reverse transcripts  $\sim 220$  nt long ( $\alpha_1$ ) and a minor band corresponding to reverse transcripts  $\sim 10$  nt shorter ( $\alpha_2$ ) were seen (Fig. 3A, lane 1). The 3' end of the  $\sim 220$ -nt reverse transcript corresponds in the DNA sequence to a position  $\sim 70$  base pairs (bp) upstream of the DHFR gene initiator codon (Fig. 2A), where the major transcription initiation site for the DHFR gene of HeLa cells has been recently mapped (11); this result therefore indicated a similarly positioned initiation site for the DHFR mRNAs from 6A3 cells.

A longer exposure of the same autoradiogram (Fig. 3A, lane 2) revealed another minor band ( $\beta$ ) corresponding to a size of  $\sim 620$  nt (with a fainter, slightly faster moving band corresponding to reverse transcripts  $\sim 25$  nt shorter), as well as other bands migrating below the  $\sim 220$ -nt band. While the latter bands probably represented premature termination products of the reverse transcriptase reaction, the upper band conceivably corresponded to reverse transcripts of either DHFR gene-related RNAs of low abundance or an unrelated RNA which had hybridized with the pH84-derived primer. To distinguish between the latter two alternatives, a large-scale primer extension reaction with 10  $\mu\text{g}$  of 6A3 polysomal poly(A) $^{+}$  RNA was carried out to isolate component  $\beta$  from an agarose–10 mM CH $_3$ HgOH gel. This component was used in an S1 protection experiment with 6A3 polysomal poly(A) $^{+}$  RNA. This analysis revealed a major protected segment corresponding to the  $\sim 220$ -nt band ( $\alpha_1$ ), resulting from hybridization of  $\beta$  with the highly predominant DHFR mRNAs (Fig. 2B), and a minor segment corresponding to the  $\alpha_2$  band detected in the primer extension experiments (cf. lanes 2 and 3 of Fig. 3B); these results therefore indicated that component  $\beta$  had been reverse transcribed from a DHFR gene-related RNA.

**Detailed mapping of the major 5' ends of the human DHFR specific RNAs.** A combination of S1 protection and primer extension analysis was carried out to map more precisely the major 5' ends of the human DHFR-specific RNAs, using the 199-nt *Hpa*II–*Bst*NI fragment and the 99-nt *Ban*II–*Bst*NI fragment of pBH31R1.8 (Fig. 1B; also see reference 35). These probes were expected to produce extended primers and S1 protected fragments 74 nt shorter than those observed with the pH84-derived 52R-Hi primer, as diagrammed in Fig. 2C and D. The actual sizes of the fragments obtained with these probes, as determined by electrophoresis through a 20-cm 7 M urea–5% polyacrylamide gel, corresponded to approximately 130, 140, and 540 nt, i.e., the sizes expected for extended primers and S1 protected fragments having the same 3' ends as  $\alpha_2$ ,  $\alpha_1$ , and  $\beta$ , respectively (data not shown). (For simplicity, these shortened forms of  $\alpha_1$ ,  $\alpha_2$ , and  $\beta$  also will be referred to as  $\alpha_1$ ,  $\alpha_2$ , and  $\beta$ .) Figure 4A shows the extended primers and S1 protected fragments from the same experiment, separated on an 80-cm sequencing gel in parallel with the products of the sequencing reactions carried out on the 199-nt fragment. Both the S1 protection products (lanes 2 and 4) and the primer extension products (lanes 1 and 3) exhibit a pronounced band representing a fragment which falls with its 3' end slightly behind the G at position  $-72$  upstream of the human DHFR reading frame (Fig. 5; references 11 and 35). This fragment, therefore, corresponds to  $\alpha_1$  in Fig. 3. Since the sequencing reaction products lack the 3' base by which they are defined and, furthermore, contain both 5' and 3' phosphates, in contrast to the extended primer and the S1-resistant fragment which have 3' hydroxyls, the alignment of the se-

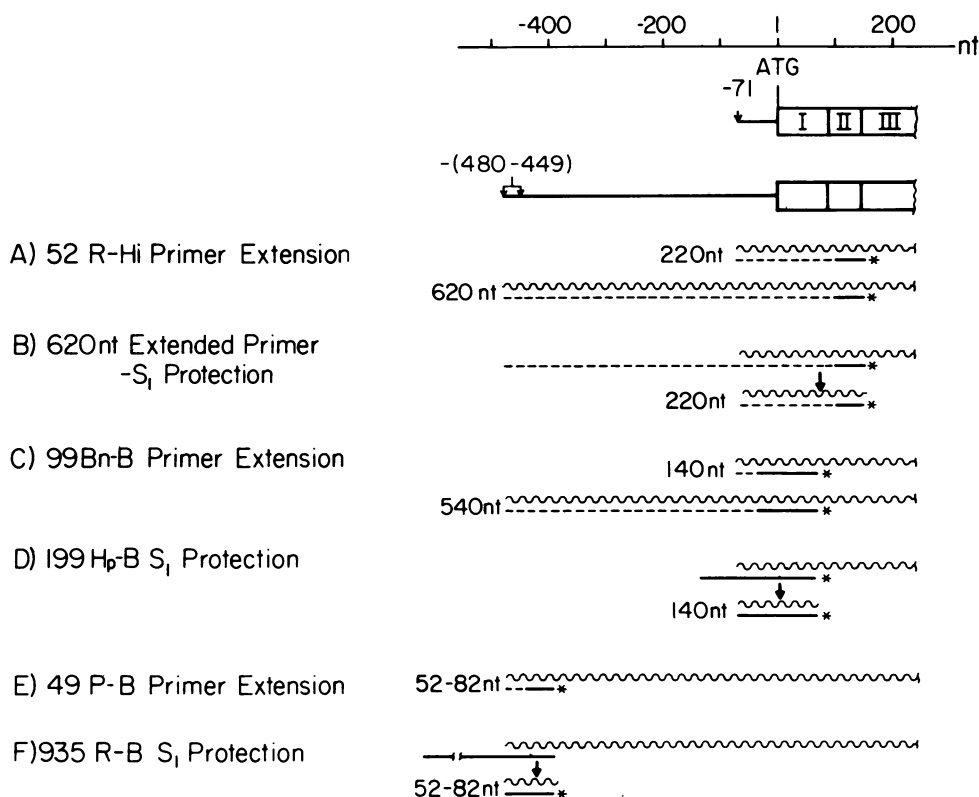


FIG. 2. Diagrammatic representation of the primer extension and S1 protection experiments using the fragments shown in Fig. 1. The transcripts to which the probes hybridize and the sizes of the extended primers and S1-resistant products obtained are shown. The scale at the top represents the nt numbering of the human DHFR reading frame and upstream region. The DHFR transcripts corresponding to the major and minor 5' ends are diagrammed below. Wavy lines represent RNA molecules, solid lines represent the DNA strands (with a star indicating the labeled end), and dashed lines represent the extended primers.

quence with these fragments is off by 1.5 nt (33). Therefore, the 3' ends of the extended primer and of the S1-resistant fragment map in correspondence to the T at position -71 upstream of the human DHFR reading frame (Fig. 5). The band corresponding to a fragment 1 nt longer than  $\alpha_1$  in Fig. 4, lane 3, could conceivably derive from reverse transcription of a minor portion of the DHFR mRNAs, starting 1 nt upstream of the major initiation point; however, the lack of a significant S1-resistant fragment of corresponding size would argue against it and, rather, speak in favor of premature termination of longer reverse transcripts. For the same reason, the bands corresponding to fragments migrating below  $\alpha_1$ , of which those 2, 3, and 9 nt shorter than  $\alpha_1$  are particularly strong, probably represent premature termination points of the reverse transcriptase reaction, rather than reverse transcripts of RNAs starting downstream of the main initiation point. One exception may be the band corresponding to  $\alpha_2$  in Fig. 3, which represents a fragment 9 nt shorter than  $\alpha_1$ . There is, in fact, an S1-resistant fragment moving one nucleotide behind this extended primer. It seems likely, however, that the S1-resistant band corresponding to  $\alpha_2$  is due to S1 nibbling of RNA-DNA hybrids in the very AT-rich 7-nt stretch starting at the -71 position (Fig. 5).

**Detailed mapping of the minor 5' ends of the human DHFR-specific RNAs.** The same analysis as that described above was carried out to map the minor, upstream 5' ends of the human DHFR-specific RNAs, using the 935-nt *EcoRI*-*Bst*NI fragment and the 49-nt *PvuII*-*Bst*NI fragment of pBH31R1.8 (Fig. 1B) for S1 protection and primer extension

experiments, respectively (diagrammed in Fig. 2E and F). When the products of these reactions were separated on a short denaturing polyacrylamide gel (Fig. 4B), a series of six major bands were observed both in the lane containing the extended primers (lane 1) and in that containing the S1-resistant fragments (lane 2). These bands, which are labeled  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$ ,  $\beta_5$ , and  $\beta_6$ , represent fragments with sizes of approximately 52, 60, 64, 70, 74, and 82 nt and correspond to band  $\beta$  in Fig. 3. The bands of the extended primers align well with the bands of S1-resistant fragments, except in the case of the  $\beta_6$  S1-resistant band which is slightly slower moving than the  $\beta_6$  extended primer band. The range of sizes of the fragments is reasonably close to that expected from the use of restriction fragments which should give extended primers and S1-resistant products 530 nt shorter than those obtained when the pHD84-derived 52R-Hi primer is used (Fig. 2). Separation of the extended primers and S1-resistant fragments from the same experiment on a thin sequencing gel, in parallel with the products of the sequencing reactions carried out on the 935-nt fragment (Fig. 4C), showed groups of bands, which, on the basis of their relative distances and intensities, clearly corresponded to  $\beta_1$  to  $\beta_6$ . The major band in each group of extended primers aligns well with the major band in the corresponding group of S1-resistant fragments, except in the case of  $\beta_6$ , where two predominant S1-resistant fragments are 1 or 2 nt longer than the major extended primer, in agreement with the results of Fig. 4B. It seems possible that the microheterogeneity of extended primers and S1-resistant fragments within each  $\beta$  band reflects a

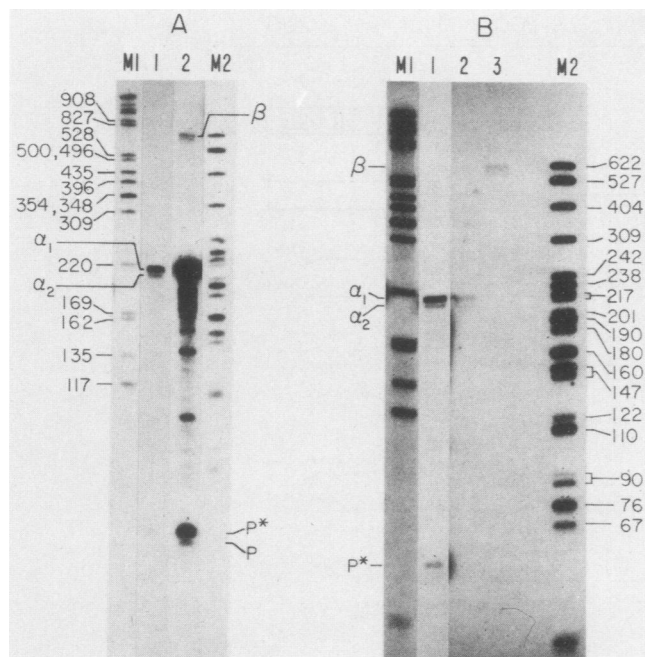


FIG. 3. Detection of minor DHFR-specific transcripts mapping with their 5' ends upstream of the major transcripts. (A) Lane 1: 1  $\mu$ g of 6A3 cell polysomal poly(A)<sup>+</sup> RNA was reverse transcribed with the pHD84-derived 52R-Hi primer, and the products were electrophoresed through a denaturing 5% polyacrylamide gel. Lane 2 is a longer exposure of lane 1 to show a slower migrating band designated  $\beta$ . P indicates the unextended primer, and P\* marks the portion of the primer which reannealed with its complementary strand and was extended 4 nt's by reverse transcriptase. M1 shows human mtDNA, and M2 shows pBR322 DNA, both digested with *Hpa*II and 3'-end labeled with [ $\alpha$ -<sup>32</sup>P]dCTP and [ $\alpha$ -<sup>32</sup>P]dGTP and the large fragment of *Escherichia coli* DNA polymerase I. (B) Ten micrograms of 6A3 cell polysomal poly(A)<sup>+</sup> RNA was used in a primer extension reaction with the 52R-Hi primer; after fractionation of the products through a denaturing agarose gel, the material corresponding to band  $\beta$  was eluted, and one-half was used in an S1 protection experiment with 6A3 RNA (75 U of S1 at 37°C). The S1-resistant products were then electrophoresed through a denaturing polyacrylamide gel (lane 2), in parallel with a sample of the products of the primer extension reaction (lane 1) and with the other half of the isolated  $\beta$  fragment (lane 3). M1 and M2 contain the same markers as in (A).

heterogeneity of start sites. A comparison with the DNA sequence in Fig. 4C shows that the major S1 resistant fragments in the groups of bands corresponding to  $\beta_1$  to  $\beta_6$  correspond with their 3' ends to nt's at positions -449, -456, -460, -466, -470, and -480 upstream of the human DHFR reading frame (again correcting for the 1.5-nt difference, as described above) (Fig. 5).

**Discrete DHFR-specific polysomal poly(A)<sup>+</sup> RNA species originate from upstream transcription initiation sites.** To define the size distribution of upstream initiating DHFR-specific transcripts, 6A3 cell polysomal poly(A)<sup>+</sup> RNA was fractionated on a denaturing agarose gel, transferred to a Zeta Probe filter, and hybridized with DHFR-specific probes. When pHD84 was used as a probe, the three previously identified DHFR mRNAs of sizes 3.8, 1.0, and 0.8 kb were observed (Fig. 6A). When the 205 B-B fragment, which maps upstream of the main transcription initiation site (Fig. 1B), was used as a probe, five major RNA species with sizes of 4.3, 3.8, 3.1, 2.1, and 1.0 kb and four minor ones with

sizes of approximately 7.3, 5.0, 1.4, and 0.8 kb were seen (Fig. 6A). (These RNA species are designated with the prefix  $\mu$  [upstream] to distinguish them from the main DHFR mRNAs.) A closer examination of the sample probed with pHD84 revealed very faint bands corresponding to some of the RNA species which reacted with 205 B-B. When the poly(A)<sup>+</sup> RNA was hybridized to the 1681 S-S probe (Fig. 1B), the same RNA species were seen as those which reacted with the 205 B-B probe (Fig. 6B); however, the 3.8, 1.0, and 0.8-kb bands gave a stronger signal with the 1681 S-S probe than with the 205 B-B probe (Fig. 6B). We interpret these results to indicate that the 3.8-, 1.0-, and 0.8-kb bands contain at least two different classes of RNA. In each of these bands the major class is represented by molecules which initiate at the -71 nucleotide position and hybridize with the 1681 S-S probe but not the 205 B-B probe; the minor class is represented by molecules which initiate at the upstream site(s) and hybridize with both probes.

The results described above strongly suggested that the discrete RNA species detected in the polysomal fraction which reacted with the 205 B-B probe mapped with their 5' ends at the upstream sites identified in this work. This interpretation has been directly confirmed by S1 protection experiments carried out with RNA samples eluted from 0.5-cm slices of an agarose gel on which 50  $\mu$ g of 6A3 poly(A)<sup>+</sup> RNA had been fractionated (data not shown).

Experiments similar to those illustrated in Fig. 6A and B, carried out with polysomal poly(A)<sup>+</sup> RNA from VA<sub>2</sub>-B and HeLa cells, revealed the same set of discrete RNA species reacting with the 205 B-B probe which were observed in the 6A3 RNA, with the exception of the 2.1- and 3.1-kb RNA species (however, a hint of the latter two species could be seen in the VA<sub>2</sub>-B pattern) (Fig. 6C). Furthermore, the species which were present in all three cell lines were increased in amount in VA<sub>2</sub>-B and 6A3 cells, relative to HeLa cells, in approximately the same proportion as the 3.8-, 1.0-, and 0.8-kb DHFR mRNAs (25, 27).

A primer extension analysis of nuclear RNA from 6A3 cells, with the 52-nt *Eco*RI-*Hinf*I fragment of pHD84 as a primer (Fig. 1A), showed that the upstream 5' ends also are present in nuclear RNA, primarily in the poly(A)<sup>+</sup> fraction; the upstream initiating RNA species represent 11% of the RNA species exhibiting the major 5' end in the nuclear fraction (data not shown), whereas this ratio is about 1% in the polysomal RNA fraction (as determined by measuring the radioactivity in the  $\alpha$  and  $\beta$  bands in an experiment similar to that of Fig. 3A).

## DISCUSSION

In the present work, by S1 protection and primer extension experiments, we have shown that the majority of the DHFR gene transcripts (~99%), including the three predominant DHFR mRNAs with sizes of 3.8, 1.0, and 0.8 kb (27), map with their 5' end at the recently identified main transcription initiation site (11) at position -71 in the 5'-flanking region of the human DHFR gene (35). Thirty nt's upstream of this site, at positions -101 to -94, there exists a possible TATA-like sequence (M. Goldberg, Ph.D. thesis, Stanford University, Stanford, Calif. 1979), CACAAATA (Fig. 5). This sequence lies within and near the 3' end of the 45-bp stretch that is homologous to the fourfold 45 to 48-bp repeat present in the mouse gene 5'-noncoding region (12). Three of the four mouse repeats contain the same TATA-like sequence, and it has been suggested that this fourfold repeat contains a DHFR mRNA 5' end (12). Flanking the proposed TATA-like sequence on both sides are almost perfect direct

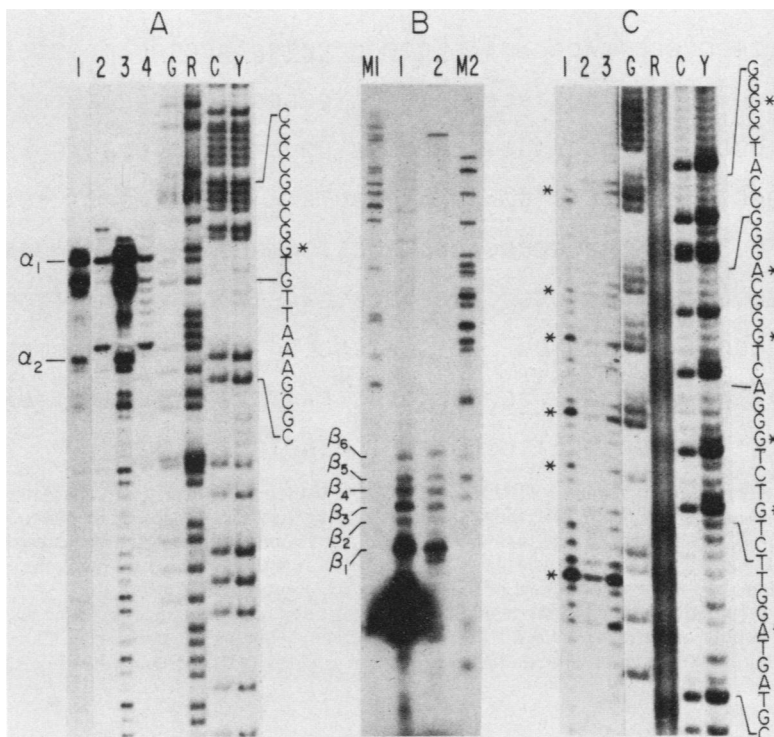


FIG. 4. Detailed mapping of the major and minor 5' ends of DHFR-specific RNAs. (A) Electrophoresis through an 80-cm sequencing gel of the primer extension and S1-resistant products obtained with 6A3 cell polysomal poly(A)<sup>+</sup> RNA and the 99- and 199-nt probes, respectively (Fig. 1B). Lanes: 1, extended primers produced with the 99-nt fragment as a primer; 2, S1-resistant products obtained with the 199-nt fragment with 80 U of S1 nuclease at 20°C; 3, long exposure of lane 1; 4, S1-resistant products obtained with the 199-nt fragment with 80 U of S1 at 37°C. G, R, C, and Y represent the products of the sequencing reactions (G, G+A, C, and C+T) carried out on the 199-nt fragment. The sequence determined is shown, with an asterisk indicating the position of the 5' end corresponding to  $\alpha_1$ . (B) The S1 protection and primer extension products from the experiments shown in panel B were electrophoresed through a 20-cm denaturing polyacrylamide gel. Lanes: 1, products of the primer extension reaction with the 49-nt probe; 2, S1-resistant products obtained with the 935-nt fragment with 200 U of S1 nuclease at 37°C. M1 and M2 contain the same markers as those in Fig. 3. (C) Samples of the S1 protection and primer extension products from the experiments shown in panel B were electrophoresed through an 80-cm sequencing gel. Lanes: 1, portion of the sample analyzed in panel B, lane 1; 2, S1-resistant products obtained with the 935-nt fragment with 100 U of S1 at 37°C; 3, portion of the sample analyzed in panel B, lane 2. G, R, C, and Y represent the products of the sequencing reactions (G, G+A, C, and C+T) carried out on the 935-nt fragment. The sequence is shown, with asterisks indicating the positions of the 5' ends corresponding to  $\beta_1$  to  $\beta_6$ .

repeats of a purine-rich sequence at nt positions -127 to -111 (GGGGCGGGGGGGCGGGG) and positions -93 to -77 (GGGACGAGGGGGCGGGG). The presence of homopolymer purine or homopolymer pyrimidine tracts has been shown to correlate with S1 hypersensitive sites near the 5' ends of transcribed genes (21, 31), and it has been suggested that they play a role in transcription initiation of eucaryotic genes. It is possible that the purine-rich sequences flanking the TATA-like sequence of the human DHFR gene have a similar role.

The most significant observation made in the present work is the identification of at least six minor 5' ends of human DHFR-specific transcripts mapping at positions -480, -470, -466, -460, -456, and -449 (11, 35); these minor 5' ends pertain to approximately 1% of the human DHFR gene transcripts found in the polysomal poly(A)<sup>+</sup> RNA. The most upstream of these sites (position -480) maps to the 3' side of a sequence (CCCCGCCCCCCAGCCCC, positions -497 to -480), which is an almost perfect inverted repeat of the purine-rich sequence flanking the major 5' end at positions -93 to -77 (Fig. 5). Again, it is interesting to speculate that this pyrimidine-rich sequence may play a role in the transcription initiation of these upstream sequences. In contrast

to the major 5' end at position -71, no TATA-like sequence is found upstream of any of the six minor 5' ends discussed here.

Transcripts of relatively low abundance have previously been found to initiate at various positions upstream (170 to 4,500 bp) of the major transcription initiation site for the mouse  $\beta$ -globin gene (17) and for the human  $\beta$ -globin gene (8, 15),  $\gamma$ -globin gene (15), and  $\epsilon$ -globin gene (1, 2). The upstream initiating transcripts of the human  $\beta$ -globin gene have been shown to be synthesized by RNA polymerase III (Pol III [8]); furthermore, at least one  $\epsilon$ -globin transcript originates in an Alu repeat (1), raising the possibility that it also is produced by Pol III, since Pol III can initiate transcription at Alu repeat sequences in vitro (13, 14, 16, 23, 28). Besides the above-mentioned observations on the globin gene system and a few instances of developmentally or metabolically regulated choice of alternative sites of transcription initiation (9, 37), the present observations represent the only other case of well-documented transcription initiation at distant upstream sites, relative to the major cap site, in endogenous cellular genes. Other previously reported instances of this phenomenon concerned cellular transformants, where the normal control of tran-



FIG. 5. Alignment of the major and minor 5' ends of DHFR specific RNAs with the sequence upstream of the human DHFR reading frame. The sequence is numbered according to Yang et al. (35); the sequence preceding the *Pst*I site (position -394 to -389) is from Chen et al. (11). The major 5' end at position -71 is indicated by the arrow, and the minor 5' ends are indicated by asterisks. The TATA-like sequence beginning at position -101 is indicated by a bracket over the sequence. The G-rich direct repeats at positions -127 to -111 and -93 to -77 are underlined with a dashed line, and an inverted repeat of the sequence from -93 to -77, at position -497 to -480, is underlined with a solid line. The DHFR translation initiation codon is boxed; another ATG in the upstream sequence, in frame with the DHFR coding sequence (at position -225 to -223), is doubly underlined. A third ATG in the upstream sequence (at position -239 to -237), not in frame with the DHFR coding sequence, as well as a termination codon (TAG, at position -95 to -93), in frame with this ATG, are singly underlined.

scription initiation of the exogenous gene was probably altered (2).

Our data show no evidence for any RNA-splicing events in the region between positions -480 and -71 and suggest

that the transcripts starting at the upstream initiation sites are spliced identically to the major DHFR-specific RNAs, at least up to the 5'-end proximal portion of the third DHFR exon. At this time, we do not know where the 3' ends of the upstream initiated RNAs map; therefore, we cannot say whether these upstream initiated transcripts contain the complete DHFR coding sequence. The observation of the presence in the polysomal poly(A)<sup>+</sup> RNA of several discrete RNA species carrying sequences mapping upstream of the main transcription initiation site strongly suggests that a substantial portion of the upstream initiating transcripts investigated here are utilized as mRNAs. However, no direct correlation has been established as yet between the six upstream 5' ends identified in this work and the individual upstream initiating discrete poly(A)<sup>+</sup> RNA species detected in the polysomal fraction.

Two translation initiation codons are present in the region between the upstream 5' ends and the major 5' end of the human DHFR-specific transcripts, one at position -239 to -237 and the other at position -225 to -223. The initiating codon at position -239 to -237 (singly underlined in Fig. 5) is not in frame with the DHFR coding sequence, and the corresponding reading frame, potentially coding for a polypeptide of 48 amino acids, terminates at the TAG at position -95 to -93. The other initiation codon, at position -225 to -223 (doubly underlined in Fig. 5), is in frame with the DHFR coding sequence, and its utilization could add 75 amino acids to the N terminus of the DHFR polypeptide. A third translation initiation codon is found at position -478 to -476, 2 nt's down from the most upstream of the minor 5' ends detected in this work. Its significance is doubtful, in view of the lack of an appreciable 5'-noncoding region in the putative corresponding transcript(s) and of the low abundance of this transcript; however, if this initiating codon were utilized, a peptide of 50 amino acids would be made, terminating at the TAG at position -328 to -326.

An interesting observation is that two of the discrete upstream initiating species found in the polysomal fraction of the methotrexate-resistant cell line 6A3 were not present,

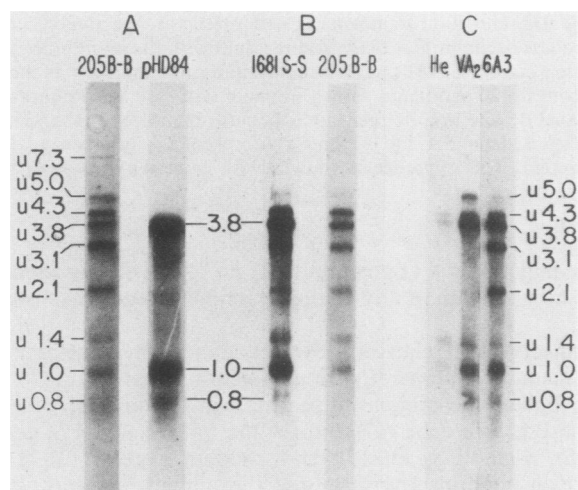


FIG. 6. Detection of discrete, upstream initiating DHFR-specific transcripts in polysomal poly(A)<sup>+</sup> RNA from 6A3, VA<sub>2</sub>-B, and HeLa cells. (A) One-microgram samples of 6A3 cell polysomal poly(A)<sup>+</sup> RNA were electrophoresed through a denaturing gel, transferred to a Zeta Probe filter, and hybridized with either the 205 B-B or the pHD84 nick-translated probe. The previously identified DHFR mRNAs with sizes of 3.8, 1.0, and 0.8 kb are indicated. The numbers preceded by a u represent the sizes of the upstream initiating DHFR-specific transcripts. (B) One-microgram samples of 6A3 cell polysomal poly(A)<sup>+</sup> RNA, treated as in (A), were hybridized with either the 1681 S-S or the 205 B-B nick-translated probe. (C) Two micrograms of polysomal poly(A)<sup>+</sup> RNA from 6A3, VA<sub>2</sub>-B, and HeLa cells were treated as in (A) and hybridized with the 205 B-B probe. The 6A3 lane was exposed for 1 day, and the VA<sub>2</sub>-B and HeLa lanes were exposed for 10 days.



or were strongly reduced, in its parental line, VA<sub>2</sub>-B, and in HeLa cells. Whether and how this phenomenon is related to the DHFR gene amplification which has occurred in 6A3 cells remain to be determined. Another significant finding is that the other discrete upstream initiating RNA species of the polysomal fraction are increased in amount, relative to the levels observed in HeLa cells, not only in 6A3 cells but also in VA<sub>2</sub>-B cells, a cell line which has undergone a moderate DHFR gene amplification in the absence of methotrexate selection (25): in both cases, the increase in the amounts of the upstream initiating transcripts is similar to that observed for the main DHFR mRNAs.

#### ACKNOWLEDGMENTS

The investigations reported here were supported by Public Health Service grants GM11726 and T32 GM07616 from the National Institutes of Health and by the Earle C. Anthony Graduate Fellowship.

The expert assistance of Caroline Anderson in some of the experiments reported here is gratefully acknowledged. The nuclear RNA utilized in this work was a generous gift of J. K. Yang. We are grateful to Benneta Keeley and Arger Drew for valuable technical help.

#### LITERATURE CITED

- Allan, M., W. G. Lanyon, and J. Paul. 1983. Multiple origins of transcription in the 4.5 kb upstream of the  $\epsilon$ -globin gene. *Cell* 35:187-197.
- Allan, M., J. Zhu, P. Montague, and J. Paul. 1984. Differential response of multiple  $\epsilon$ -globin cap sites to *cis*- and *trans*-acting controls. *Cell* 38:399-407.
- Amaldi, F., and G. Attardi. 1968. Partial sequence analysis of ribosomal RNA from HeLa cells. I. Oligonucleotide pattern of 28S and 18S RNA after pancreatic RNase digestion. *J. Mol. Biol.* 33:737-755.
- Amalric, F., C. Merkel, R. Gelfand, and G. Attardi. 1978. Fractionation of mitochondrial RNA from HeLa cells by high resolution electrophoresis under denaturing conditions. *J. Mol. Biol.* 122:301-319.
- Auffray, C., and F. Rougeon. 1980. Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. *Eur. J. Biochem.* 107:303-314.
- Bailey, J. M., and N. Davidson. 1976. Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. *Anal. Biochem.* 70:75-85.
- Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7:1513-1523.
- Carlson, D. P., and J. Ross. 1983. Human  $\beta$ -globin promoter and coding sequences transcribed by RNA polymerase III. *Cell* 34:857-864.
- Carlson, M., and D. Botstein. 1982. Two differently regulated mRNAs with different 5' ends encode secreted and intracellular forms of yeast invertase. *Cell* 28:145-154.
- Carothers, A. M., G. Urlaub, and L. A. Chasin. 1983. Structure of the dihydrofolate reductase gene in Chinese hamster ovary cells. *Nucleic Acids Res.* 11:1997-2012.
- Chen, M.-J., T. Shimada, A. D. Moulton, A. Cline, R. K. Humphries, J. Maizel, and A. W. Neinhuis. 1984. The functional human dihydrofolate reductase gene. *J. Biol. Chem.* 254:3933-3943.
- Crouse, G. F., C. C. Simonsen, R. N. McEwan, and R. T. Shimke. 1982. Structure of amplified normal and variant dihydrofolate reductase genes in the mouse sarcoma S180 cells. *J. Biol. Chem.* 257:7887-7897.
- Di Segni, G., G. Carrara, G. R. Tocchini-Valentini, C. C. Shoulters, and F. E. Baralle. 1981. Selective *in vitro* transcription of one of the Alu family repeats in the 5' flanking region of the human  $\epsilon$ -globin gene. *Nucleic Acids Res.* 9:6709-6722.
- Duncan, C., P. A. Biro, P. V. Choudary, J. T. Elder, R. R. C. Wang, B. G. Forget, J. K. DeRiel, and S. M. Weissman. 1979. RNA polymerase III transcriptional units are interspersed among human non- $\alpha$ -globin genes. *Proc. Natl. Acad. Sci. U.S.A.* 76:5095-5099.
- Grindlay, G. J., W. G. Lanyon, M. Allan, and J. Paul. 1984. Alternative sites of transcription initiation upstream of the canonical cap site in human  $\gamma$ -globin and  $\beta$ -globin genes. *Nucleic Acids Res.* 12:1811-1820.
- Haynes, S. R., and W. R. Jelinek. 1981. Low molecular weight RNAs transcribed *in vitro* by RNA polymerase III from Alu-like dispersed repeats in Chinese hamster DNA are also found *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 78:6130-6134.
- Hofer, E., and J. E. Darnell. 1981. The primary transcription unit of the mouse  $\beta$ -major globin gene. *Cell* 23:585-593.
- Holland, L. J., and L. J. Wagh. 1983. Efficient recovery of functionally intact mRNA from agarose gels via transfer to an ion-exchange membrane. *Nucleic Acids Res.* 11:3283-3300.
- Hoopes, B. C., and W. R. McClure. 1981. Studies on the selectivity of DNA precipitation by spermine. *Nucleic Acids Res.* 9:5493-5504.
- Jeanteur, P., F. Amaldi, and G. Attardi. 1968. Partial sequence analysis of ribosomal RNA from HeLa cells. II. Evidence for sequences of nonribosomal type in 45S and 32S ribosomal RNA precursors. *J. Mol. Biol.* 33:757-775.
- Larsen, A., and H. Weintraub. 1982. An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin. *Cell* 29:609-622.
- Lewis, J. A., J. P. Davide, and P. W. Melera. 1982. Selective amplification of polymorphic dihydrofolate reductase gene loci in Chinese hamster lung cells. *Proc. Natl. Acad. Sci. U.S.A.* 79:6961-6965.
- Manley, J. L., and M. T. Colozzo. 1982. Synthesis *in vitro* of an exceptionally long RNA transcript promoted by an AluI sequence. *Nature (London)* 300:376-379.
- Masters, J. N., B. Keeley, H. Gay, and G. Attardi. 1982. Variable content of double minute chromosomes is not correlated with degree of instability in methotrexate-resistant human cell lines. *Mol. Cell. Biol.* 2:498-507.
- Masters, J. N., J. K. Yang, A. Cellini, and G. Attardi. 1983. A human dihydrofolate reductase pseudogene and its relationship to the multiple forms of specific messenger RNA. *J. Mol. Biol.* 167:23-36.
- Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499-560.
- Morandi, C., J. Masters, M. Mottes, and G. Attardi. 1982. Multiple forms of human dihydrofolate reductase messenger RNA; cloning and expression in *E. coli* of their DNA coding sequences. *J. Mol. Biol.* 156:583-607.
- Pan, J., J. T. Elder, C. H. Duncan, and S. M. Weissman. 1981. Structural analysis of interspersed repetitive polymerase III transcription units in human DNA. *Nucleic Acids Res.* 9:1151-1170.
- Pontén, J., F. Jensen, and H. Koprowski. 1963. Morphological and virological investigation of human tissue cultures transformed with SV40. *J. Cell. Comp. Phys.* 61:145-154.
- Prentki, P., F. Karch, S. Iida, and J. Meyer. 1981. The plasmid cloning vector pBR325 contains a 482 base-pair-long inverted duplication. *Gene* 14:289-299.
- Schon, E., T. Evans, J. Welsh, and A. Efstratiadis. 1983. Conformation of promoter DNA: fine structure mapping of S1 hypersensitive sites. *Cell* 35:837-848.
- Setzer, D. R., M. McGrogan, and R. T. Shimke. 1982. Nucleotide sequence surrounding multiple polyadenylation sites in the mouse dihydrofolate reductase gene. *J. Biol. Chem.* 257:5143-5147.
- Sollner-Webb, B., and R. H. Reeder. 1979. The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in *X. laevis*. *Cell* 18:485-499.
- Weiss, M. C., B. Ephrussi, and L. J. Scaletta. 1968. Loss of T-antigen from somatic hybrids between mouse cells and SV40-transformed human cells. *Proc. Natl. Acad. Sci. U.S.A.*

- 59:1132-1135.
35. Yang, J. K., J. N. Masters, and G. Attardi. 1984. Human dihydrofolate reductase gene organization: extensive conservation of the (GC)-rich 5' noncoding sequence and strong intron size divergence from homologous mammalian cells. *J. Mol. Biol.* **176**:169-187.
36. Yang, R. C.-A., J. Lis, and R. Wu. 1979. Elution of DNA from agarose gels after electrophoresis. *Methods Enzymol.* **68**:176-182.
37. Young, R. A., O. Hagenbückle, and U. Schibler. 1981. A single mouse  $\alpha$ -amylase gene specifies two different tissue-specific mRNAs. *Cell* **23**:451-458.